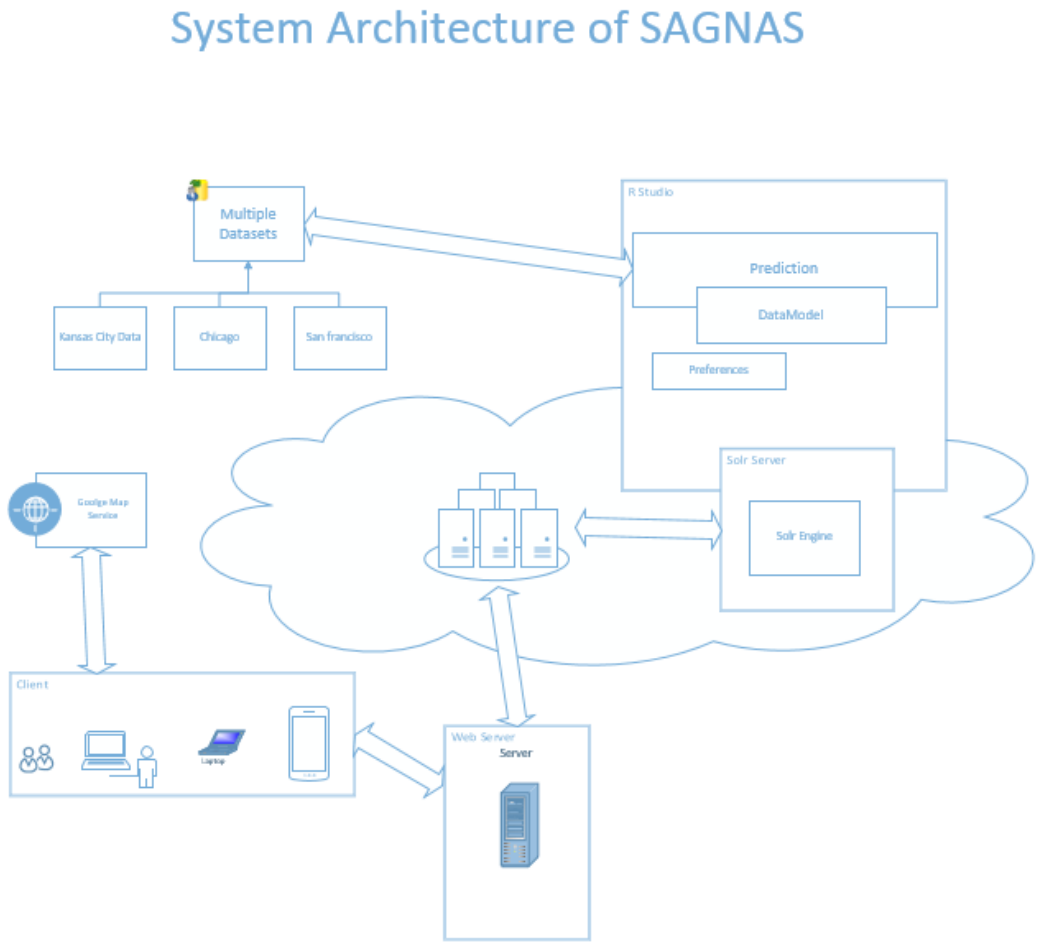
SAGNAS

Increment 3 – Group 9

Framework Specification:

System Architecture Diagram:

This system is a cloud based mobile application where the user can see the future of crime in a specific region. The Architectural components are Mobile Client, Webserver, Solr(Indexing Engine), R for statistical analysis.



Domain Model:

- Data Sources:

We have datasets from different cities with the data updated almost every two days. Each dataset contains data from 2001 to present.
The cities include Kansas City, Chicago, and San Francisco.

- Methodology and Algorithm:
  - Before designing a model we need to decide upon on what level of data has to be used to describe the model. So we decided to go with month level i.e. during which period the crime took place and based on crime type.
  - The datasets chosen has several records and each record contain attributes like location, date, time, type of crime, beat, Gender, Involvement, etc.
  - Based on the year we filter the event records and generated dataset subsets.
  - To increase the accuracy of the prediction we choose only past 5years because the safety measures from 2000 has improved to a greater extent and considering the records from 2001 won't provide that accurate results.
  - Once we have the data records filtered based on year, we further divide the dataset to months in each year which helps to find a pattern in the crime rate that shows exactly during which season the crime is happening the most.
  - Further the dataset is divided in terms of crime type where we selected only major categories for simplicity.
  - Get the total number of crimes happening in a specific month and calculate the crime rate in a region, i.e. with respect to population in a region.
  - The algorithm used is aggregation along with clustering.
- **Analytic Tools:**
  - 'R' a powerful statistical tool is used to simulate the proposed model which has inbuilt algorithms and where we can generate the required subsets easily.
- **Analytical Tasks:**
  - The analytical task is to predict the rate of crime on a specific region and show the region is safe or not based on the following table

Grouped To	Rate
Relax	< 5%
Safe	6% - 15%
Moderate	15% - 25%
Less Dangerous	25% - 35%
Dangerous	> 36%

Application Specification:

- **Features List:**
  - A web service to call solr and send the data to the client.
  - Using google web-service to load the maps.

    - A feature that displays zip code regions on a map.
    - A feature that marks the region to show the user that it is safe or not.
- **Mobile Client:**
    - The mobile client is a web-application which is implemented using HTML5, CSS3, JQuery and Twitter Bootstrap.

Implementation

- The system stores all the data SOLR prior to which the data is modeled using 'R'.
- The implementation tasks are as follows:
    - Web Service that requests solr w.r.t zip code and updates the UI, which is filled with google map.
    - On Selection of specific region the detailed information is displayed i.e. the percentage of the crime rate, no of crimes in a month in an average, etc.

Fields stored in SOLR:

- id
- Crime_Total_i
- Assault_Total_i
- Theft_Total_i
- April_Crime_Total_i
- Assault_April_i
- Theft_April_i
- April_Assault_Crime_Rate_s
- April_Theft_Crime_Rate_s
- April_Crime_Rate__From_Year_s